# Tough Behavior in the Repeated Bargaining Game. A Computer Simulation Study.

Linh Chi Nguyen[*]
University of Trento
Via Inama 5
Trento, Italy
linhchi.nguyen@unitn.it

Luciano Andreozzi
University of Trento
Via Inama 5
Trento, Italy
luciano.andreozzi@unitn.it

## ABSTRACT
Bargaining behavior occupies an important part in economics literature, or social sciences in general. Although there is an extensive simulation literature on social tradeoff in the Prisoner's Dilemma and the one-shot bargaining game, little has been done for the repeated bargaining game. Part of reason for this neglect is that, despite having a continuum of Nash equilibria, under homogeneous settings, the one shot bargaining game consistently gives a stable equilibrium of fairness (50-50 division), robust to many kind of tough perturbations. However, it's true that social interaction doesn't always yield unconditional egalitarianism. Hence we simulate a population of homogeneous agents playing the repeated bargaining game to test the stability of the 50-50 norm under evolutionary force. It turns out that when it comes to repeated interaction, the fair norm no longer stands strong.

## Categories and Subject Descriptors
J.4 [**Social and behavioral sciences**]: Economics

## General Terms
Economics

## Keywords
Nash Demand Game, Evolutionary Game Theory, Simulation

## 1. INTRODUCTION
In economics, bargaining behavior can be the salary negotiation between the worker union and the managerial board. Or it can be about setting a price to divide the distance between willingness to pay and willingness to sell of a buyer and a seller. The simplest abstract version is in the analogy

---

of dividing a pie. Though simple, it is able to represent a vast range of actions from daily bargaining in the market between individuals to claiming natural resources among conventional nations. By studying this game, we hope to contribute useful discussion on the question why the pie is divided differently in different societies.

### 1.1 The game

| NDG | High | Medium | Low |
|--------|------|--------|------|
| High | 0,0 | 0,0 | **8,2** |
| Medium | 0,0 | **5,5** | 5,2 |
| Low | **2,8** | 2,5 | 2,2 |

**Table 1: Nash Demand Game payoff matrix**

We consider a specific version of the bargaining game called Nash Demand game (NDG) in which two players divide a fixed pie of size 10 (Table 1). Each can claim a {High, Medium, Low} fraction of the pie which is equivalent to {8,5,2} out of 10. The hypothesis is that though the repeated interaction helps stabilise the cooperative behavior, it destabilises the *fair* behavior. In the Prisoner's Dilemma (PD), as much as everyone prefers the optimal cooperative outcome, the selfish logic pushes us all in the opposite direction. However, if we play repeatedly, cooperation can emerge and the society can prosper. In a broader sense, regarding reputation building and social image, the one-shot interaction among homogeneous players are equivalent to anonymous matching. Without reputation to signal and gain trust, everybody defects everybody. It's the possibility of tomorrow that makes cooperation partly be in the interest of both players. Hence we'd like to address the similar question: Is 50-50 division still a stable norm when agents interact repeatedly?

### 1.2 Evolutionary Game Theory
The difference between classical and evolutionary game theory (EGT) [8] lies in the rationality assumption of the player. In the classical paradigm, players are capable of doing complex optimization and infinite recursive thinking of common knowledge. This is criticised for not fitting with human behavior in reality. Until now the narrative has substantially shifted from these ideal players to less demanding subjects. As behavioral and experimental economists are incorporating psychological factors and neuro-economists are scanning brains to crack open the black box of the decision maker,

---

theorists bring evolution into game theory to explain how people can reach the so-called hyper-rational solution. Descriptively, the evolutionary process (or "generalised Darwinism" [8]) is an iteratively updating loop that manifests beyond its origin field of biology. This process consists of two mechanisms: selection and mutation. The selection part (that measures on some particular dimension, such as payoff) will narrow down the set of fittest *survivors* and the mutation part is to keep feeding variety into the selection pool. Hence the equivalent to equilibrium concept here is the "robust rest point of the dynamics" [13].

**Evolutionarily Stable Strategy** The central concept in EGT is an evolutionarily stable strategy (ESS), proposed by Maynard Smith [8]. This concept is used to describe the *persistence* of one strategy against another, illuminating in a population context. The scenario is as following: there is one population hosting exclusively strategy A. A can be pure or mixed. When there is $\epsilon$ mutation B appearing in the population, A is said to be evolutionarily stable against B if A can repel B. Otherwise B invades A. Or they can be neutrally stable and coexist at any possible ratios.

**Replicator Dynamics and Adaptive Learning** What Maynard Smith provides is a static pairwise test of 2 strategies. Economics theory then develops a sharp mathematical tool to illuminate its underlying dynamics. Our simulation tries to approximate the replicator dynamics (RD) with reinforcement learning by fixed aspiration, as in Vega-Redondo [13], chapter 11. Intuitively speaking, in the RD, the replicator is the strategy and the dynamics of the replicators is a vector field representing the evolution of the population over time. The underlying learning mechanism of the dynamics is reinforcement learning. After we have the payoff vector of the whole population, the agents having the chance to learn will change their strategy according to this payoff vector. Better strategies have better chance to proliferate over time at the expense of the poor doers. Technically, the growth rate of one strategy is the difference between its average payoff and the population average payoff.

## 2. SIMULATION SETTINGS

The PD has been extensively studied in both theoretical [4] [6] [7] [10] and simulation literature [1] [9] [12], also the bargaining game and social norm [2] [3] [11] [14]. In the following we consider the repeated NDG.

## 2.1 Strategy in the Repeated NDG

In a one-shot game, there are only 3 possible strategies {High, Medium, Low}. In a repeated game, the number of strategies grows exponentially with the number of rounds per match. We model these strategies similar to the rule-based strategies in the iterated PD [10], in which the agent conditions its next behavior on the outcome of the previous round. Because the 3x3 NDG has 9 possible outcomes, a strategy will have 9 rules to specify the next move based on these 9 outcomes, plus 1 rule prescribing the first move. These rules are deterministic because it prescribes claims with probability 1. Hence there are $3^{10}$ strategies in the entire set of deterministic strategies. Figure 1 shows the representation in finite state machine of a strategy that prescribes the agent to start by playing Medium then to play best response to whatever the previous move of the oppo-

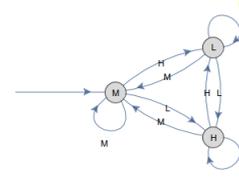nent. In Figure 2, there are 3 basic unconditional strategies that always claim High, Medium and Low no matter what.



**Figure 1: An accommodating strategy, starting to play Medium, then playing best response to the previous move of the opponent**



**Figure 2: Unconditional strategies**

## 2.2 Simulation Cycle

| Round | Accommodator | All Highs | Ac | AH |
|-------|--------------|-----------|----|----|
| 1 | M | H | 0 | 0 |
| 2 | L | H | 2 | 8 |
| 3 | L | H | 2 | 8 |
| .. | .. | .. | .. | .. |

**Table 2: Accommodator meets All Highs**

A typical simulation cycle has 3 phases. Initially, each agent adopts randomly a deterministic strategy. In the matching phase, they are randomly pair-matched to play the NDG for a number of rounds. Then we take the mean of the resulted payoff sequences to calculate the relative fitnesses. Table 2 shows the move sequences and the payoff sequences when Accommodator meets All Highs. The learning phase starts after we have the fitness vector of the whole population. A fraction of the population will be allowed to observe the population and choose to change their strategy. They use a weighted lottery that gives all available strategies a chance (relative to their fitness) to be chosen. In the mutation/mistake phase, new strategies are added into the population. A fraction of the population will be allowed to mutate from their current strategy.

## 3. RESULT: THE ONE SHOT GAME

In Figure 3, we have Ox being $p_L$ (the fraction of population playing Low), Oy being $p_M$ (the fraction of the population playing Medium). The graph a shows 3 regions of population states in which it's best to play Low, Medium and High. Figure 3b shows the theoretical RD of the one-shot NDG. There are 3 rest points of the dynamics that are interesting (marked in Figure 3a). Point 3 corresponds to the state of population in which all agents play Medium. Point 1 and 2 are mixed rest points. However, we can see that point 2 has zero basin of attraction. The basin of attraction of point 1 is small, it takes some mistakes and the population will fall into the trajectory of another state. In contrast, the basin of attraction of the Medium equilibrium is very
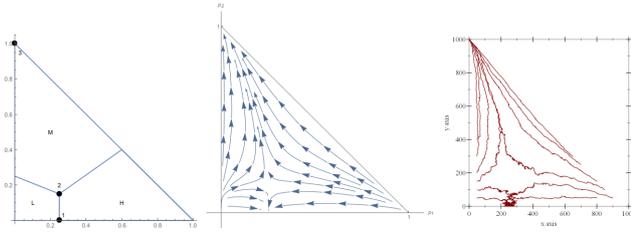
**Figure 3: a, Regions of best responses for each population state. b, RD of the one shot NDG, with respect to $pL$ and $pM$. c, The simulated RD approximating the prediction in b.**

large and nearly impossible to escape by mistaking. So the stable equilibrium is the 50-50 division (point 3). Figure 5c is the evolution of population state that we simulate, quite approximating the theoretical prediction.

## 4. RESULT: THE REPEATED GAME

Figure 4 is 20 typical runs of simulation, with different learning rate and rounds per match. We plot the population mean over cycles. Note that 1 round per match is the one shot game hence the bottom 4 plots are simulations of the one-shot game. We can see that the theoretical prediction of the one-shot game in which fairness preserves is replicated here. The surprising result is in the repeated game simulations. In general, when the rounds per match increase (along the vertical direction), there are periods that the population's payoff is slightly but significantly less than 5. Second, there are periods that the population mean goes down below 2.5. When the speed of learning increases (fixed amount of mutation) along the horizontal direction, the down periods are more likely to span bigger. We investigate the demographic of the population in those periods as follows.
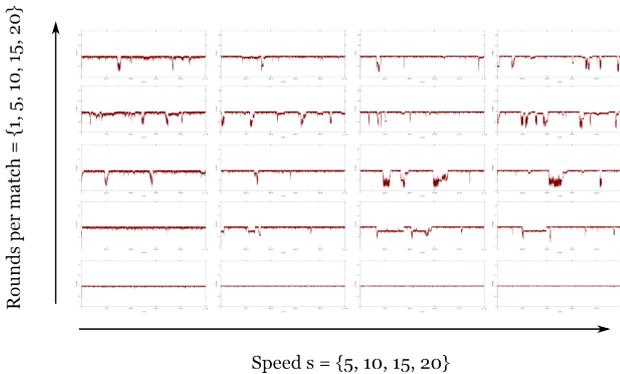


**Figure 4: Population mean over cycles, different learning rate and rounds per match. The bottom 4 are of the one-shot game, others are of the repeated.**

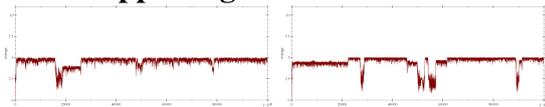## 4.1 Less Appealing Fairness



**Figure 5: Periods hosting tough fair players**

When the population mean is slightly less than 5, the population is full of tough strategies that only revert to playing 5 after the first round being aggressive. The demographic of one such period in Figure 5 is shown in Figure 6. The strategy in Figure 6a takes up 23% of the population. It immediately claims High in the first round and switches to claim Medium in the second round, independent from what the opponent does. From that second state, this strategy is a moderate accommodator. The second strategy (Figure 6b) takes up 10%. It starts by playing High also. If the opponent is also tough in the first round, it retreats to playing Medium and tries to stay in that state. Medium state is an absorbing state in this machine hence it is a Fair strategy. The third and fourth strategies (both 9% and not shown here) start with a tough claim then revert to a medium claim.



**Figure 6: Top two strategies: 23% & 10%**

Overall, the four dominant types in this population state all start off the game by playing tough but for the first round only. After that they behave fairly good. Some retreat immediately regardless of the opponent move but some keeps exploiting if the opponent keeps compromising. In this state, because no strategies start by playing Low, all of them retreat right in the second round. Hence a typical payoff sequence in this period is: 0 5 5 5 5 5... We can see that though fairness stands, it is costly because of delayed negotiation. Agents only reach the efficient (and nice) agreement from the second round onward.

## 4.2 True Inequality and Inefficient Matching

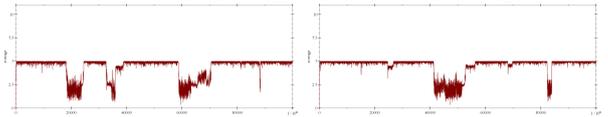### 4.2.1 A Mixture of Tough and Weak Strategies



**Figure 7: Periods with population mean very low**

There are other periods that host very unequal mixture of strategies. These periods can be recognised by very low population mean (Figure 7). Examining one cycle of low average payoff in Figure 7, we describe the population state as follows. The top two strategies (34%) are essentially All Highs. They start playing High and never leave that state no matter what. Also the third one. The fourth strategy is similar to an accommodator. It starts playing Medium and keeps being fair if the opponent does so. But it can switch to Low and High and from then. The fifth machine is a weak accommodator. Because this population state is indeed full of aggressive strategies that insist on getting 8 no matter what, the last two strategies retreat to play Low eventually. The low population mean is due to inefficient matching among the tough ones (both get 0) and among the weak ones (both get 2).
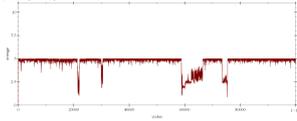
### 4.2.2 Alternators



**Figure 8: 1 run of speed 20, rounds per match 200**

Figure 8 shows another example of periods having very low population mean but with different population demographic. At cycle 65000 of Figure 9, there are 3 most popular machines. The first two machines (73%) start claiming High and move back and forth between High and Medium altogether. They both get the same payoff sequence which is an alternating sequence of 0 and 5 (hence the average is 2.5).

**Remark** As these two kinds of states is too inefficient, sooner or later the population will make it way out. However, these "bad" periods significantly exist along the evolution of the population.

## 4.3 Sufficiently Patient Agents

Up until now, we've always taken the mean of the payoff sequence to calculate the fitness. This is equivalent to the assumption that agents are infinitely patient. The payoff at round 10 is just equally important as that at round 1. To relax this, we run simulations with different discounting factor $\delta$ in calculating the payoff sequence. If $\delta$ is small, the agent is very impatient and wants the benefit to come as soon as possible. At $\delta = 0$, it's such as they are playing the one shot game. A patient agent, on the other hand, can bear the cost of initial rounds and prefers the sequence (0 0 0 8 8 8 ..) to the sequence (5 5 5 5 5 5 ..). Hence, we speculate that impatient agents will gets to the 50-50 division quite fast and patient agents will make the negotiation result messier.
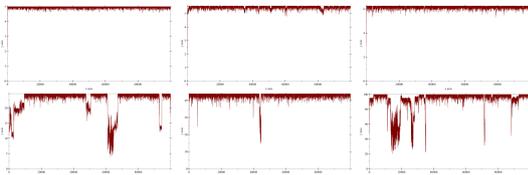


**Figure 9:** $\delta = \{0, 0.1, 0.2, 0.8, 0.9, 1\}$

Indeed, the simulations (Figure 9) show that with $\delta$ sufficiently small, the 50-50 equilibrium is very stable. However, when $\delta$ tends to 1, the "bad" periods appears consistently over simulations. Hence we note that if agents are sufficiently patient, the negotiation tends to go bad easier than in the case of myopic agents.

## 5. CONCLUSION

As stated at the beginning, our motivation is on how repeated interaction affects two behaviors: cooperative and bargaining. We speculate that, lengthening the interaction horizon is good for fostering cooperation but it may do harm to the case of resolving conflicts. Specifically in the PD, classical game theory (CGT) can sustain any outcomes that are better than the always-defecting because the agreement will be well kept under a credible punishment threat. Repetition doesn't just secure these new better points as Nash

equilibria, it makes them subgame perfect equilibria [6]. The evolutionary literature agrees that cooperation can emerge with repeated interaction but proposes that the population goes through a cycle of cooperating and defecting because no punishment is good enough to keep cooperation in place [4] [7]. In the bargaining game, CGT with rational players says that negotiation ends in the first round with efficient outcomes [11]. EGT also shows that the fair division is the more stable outcome in the one shot negotiation [3]. As our simulation adds, once the game is repeated, it takes costly mechanisms to sustain fairness. At worst, there is the dominance of tough strategies that stubbornly don't retreat because there are weak strategies accommodating them. Interpreting in term of patience, the zero-patient agent only cares about today hence the nice division is very stable in such society. As $\delta$ increases, there is a threshold that once the agents become sufficiently patient, the society will go through periods of delayed negotiation and inefficient bargaining. Note that overall, these down periods significantly drive down the population mean, but not much. The population spends a lot of time in the 50-50 norm.

## 6. REFERENCES

[1] Axelrod R.:The evolution of cooperation: Revised edition. New York: Basic Books (2006)

[2] Binmore K., Samuelson L., Young P.: Equilibrium Selection in bargaining models. Games and Economic Behavior, 45, 296-328 (2003)

[3] Binmore K., Piccione M., Samuelson L.: Evolutionary Stability in Alternating-Offers Bargaining Games. Journal of Economic Theory, 80, 257-291 (1998)

[4] Boyd R., Lorberbaum J. P.: No pure strategy is evolutionarily stable in the repeated Prisoner's Dilemma game. Nature, 327, 58-59 (1987)

[5] Hilbe C., Nowak M. A., Sigmund K.: Evolution of extortion in Iterated Prisoner's Dilemma games. PNAS, 110, 6913-6918 (2013)

[6] Kreps D. M.: A course in microeconomic theory. Princeton University Press (1990)

[7] Lorberbaum J.: No strategy is evolutionarily stable in the repeated Prisoner's Dilemma. J. theor. Biol. 168, 117-130 (1994)

[8] Maynard Smith, J.: Evolution and Theory of Games. Cambridge University Press. (1982)

[9] Nowak M., Sigmund K.: A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature, 364, 56-58 (1993)

[10] Press W. H., Dyson F. J.: Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. PNAS, 109, 10409-10413 (2012)

[11] Rubinstein A.: Perfect equilibrium in a bargaining model. Econometrica, 50, 97-109 (1982)

[12] van Veelen M., Garcia J., Rand D. G., Nowak M. A.: Direct reciprocity in structured populations. Proceedings of the National Academy of Sciences, 109, 9929-9934 (2012)

[13] Vega-Redondo, F.: Economics and Theory of Games. Cambridge University Press (2003)

[14] Young P.: Individual Strategy and Social Structure. Princeton University Press (1998)